Niveau(x) d'enseignement: 0 Numéro EDUQ : 8170860
Centre d'intérêt: 32

Duchesne, Louis, 1947-.

Modèle de flux des clientèles scolaires : rapport d'étape no 1 / Louis Duchesne, Claude St-Germain ; [pour le] Ministère de l'éducation, Direction générale de la planification, Démographie scolaire. -- [Québec] : Le Ministère, 1973.
38 f. : graph. ; 28 cm. -- (Document de travail / Service de la démographie scolaire du Québec; 9.08)

Titre de la couv.

Cote Dewey: 371.219'714/D829m Cote LC: LB/2846

Descripteur(s): *Population étudiante *Taux de scolarisation *Taux de passage *Statistique *Prévision *Analyse statistique *Banque de données

Identificateur(s) de sujet: *Modèle de flux *Coefficient de transition *Fichier informatisé

Premier rapport d'étape concernant l'élaboration d'un modèle de flux des clientèles scolaires, cette étude fait suite au Document de travail 9.07. Le modèle vise à analyser les mouvements des clientèles scolaires de tous les niveaux du système d'éducation. Il servira aux fins de prévisions de clientèles et de diplômés à court et à long terme. Ce cahier précise la méthodologie qui sera utilisée dans le calcul des flux (coefficient de transition) et ses exigences quant au type de données nécessaires. Le problème du choix des variables est examiné de même que la méthodologie des prévisions.

Trois annexes complètent le tout: 1. Description sommaire et préliminaire d'un modèle de flux des clientèles étudiantes, 2. Coefficients de transition, matrices et chaînes de Markov. Application au système scolaire, 3. Relations entre quelques indices utilisés en démographie scolaire: les taux de scolarisation, les taux de passage et les coefficients de transition.

Type de document: r
Auteur de l'analyse: ar

E D U Q 8170860

Ministère de l'Education
Direction générale de la Planification

Démographie scolaire
Document de travail

D.T. 9.08

MODELE DE FLUX DES CLIENTELES SCOLAIRES

Rapport d'étape no 1

Louis Duchesne
Claude St-Germain

Juin 1973

E D U Q 8170860

PRESENTATION

Le texte qui suit est le premier d'une série de "rapports d'étape" concernant l'élaboration d'un modèle de flux des clientèles scolaires du Québec. Il fait suite au Document de travail 9.07 "Ebauche d'un modèle de flux des clientèles scolaires".

Rappelons brièvement le but du modèle, déjà décrit dans ce document. Le modèle de flux analysera les mouvements des clientèles scolaires de tous les niveaux du système d'éducation du Québec. Il servira aux fins de prévisions de clientèles et de diplômés (finissants) à court et à long terme(1).

A cette fin l'utilisation de matrices nous semble être la méthode la plus intéressante. Le présent document tente donc de préciser la méthodologie qui sera utilisée dans le calcul des flux et ses exigences quant au type de données nécessaires. Nous examinerons en même temps le problème du choix des variables, puis nous ébaucherons la méthodologie des prévisions.

---

(1) On retrouvera à l'Annexe I un exemple de demande de prévisions, avec la période de temps visée et les ventilations désirées, à laquelle le modèle tentera de répondre.

Cadre conceptuel théorique

La population scolaire comme toutes les populations présente une structure qui évolue dans le temps grâce à des entrées, des sorties et des changements internes. Le moyen le plus raffiné de décrire une population consiste à écrire et analyser des biographies d'individus réunis en groupes significatifs. Vivre, c'est faire l'expérience de certains événements au cours de certaines étapes; vivre un événement, c'est passer d'un état à un autre.

L'outil statistique le plus utile à la description de ces flux est le coefficient de transition qui donne à une étape donnée, pour un certain état, la distribution en pour-cent des individus dans les états à l'étape suivante. En d'autres mots, le coefficient de transition nous permet de retrouver les individus à l'étape suivante. Un coefficient q_{ij} de .75 par exemple nous dit que 75% des individus en i à l'étape t se trouveront en j à l'étape t+1.

On peut construire une matrice avec l'ensemble des coefficients. Supposons qu'il y a n états; nous aurons alors une matrice n X n qui présentera l'ensemble des transitions possibles.

$$A = \begin{bmatrix} q_{11} & \cdots & q_{1n} \\ \vdots & & \vdots \\ q_{ij} & & \\ \vdots & & \vdots \\ q_{n1} & \cdots & q_{nn} \end{bmatrix}$$

Sur la diagonale on trouve les coefficients q_{ii}, soit les proportions d'individus qui restent dans le même état.

A partir d'une telle matrice, on peut calculer les indices décrivant l'intensité et le calendrier de la scolarisation et faire des projections (sans immigration). On peut connaître par exemple la proportion des individus en i à l'étape t qui seront en j à l'étape t+3..., mesurer la durée moyenne de séjour dans l'état j des individus actuellement en i... On trouvera à l'annexe II plus de renseignements concernant l'utilisation des matrices.

Exigences du modèle

L'opération du modèle tel qu'on vient de le décrire a des exigences, qu'on peut regrouper en trois catégories:

1) Définition des états de la matrice
2) Définition de la variable temps
3) Définition du type de données nécessaires au calcul des coefficients de transition.

1) Définition des états de la matrice

La définition des états de la matrice procède de deux opérations: 1) la définition des états du système scolaire en terme d'activités scolaires et de niveaux d'avancement; 2) le choix de sous-populations (ou des variables).

Définition des états du système scolaire

Le système scolaire sert de base structurale au modèle. Le parcours scolaire a déjà été décrit dans le D.T. 9.07. Ce parcours scolaire nous conduit, dans une première phase à définir un (1) état pour le niveau maternelle et sept (7) états pour le niveau élémentaire, incluant la possibilité d'une année de récupération entre l'élémentaire et le secondaire.

Au niveau secondaire, la distinction entre programme général et programme professionnel après Secondaire II nous amène à définir un minimum de dix (10) états, incluant les "états absorbants"(1) correspondant aux certificats d'études secondaires. Au collégial, la distinction entre six (6) concentrations au programme général et cinq (5) au professionnel conduit au nombre de 38 états (incluant 11 DEC.)

---

(1) Voir annexe II.

Enfin au niveau universitaire, les états retenus peuvent être définis en terme de "crédits-programmes", c'est-à-dire les niveaux d'avancement (crédits accumulés) dans chacun des programmes, classifiés selon la taxonomie Clarder. Si au 1er cycle on regroupe les programmes selon les 24 secteurs Clarder, et si on définit 4 niveaux d'avancement (incluant l'état de diplômé) on aboutit à 96 états scolaires.

On remarquera que plus la population est faible, plus on doit la diviser en un grand nombre de sous-populations. On a besoin de ventilations plus raffinées quand les données sont statistiquement les plus précaires.

Choix des sous-populations

A l'intérieur des états du système scolaire, on peut ventiler la population étudiante selon plusieurs variables. Un choix des principales variables a déjà été justifié dans le D.T. 9.07. Ce sont, mises à part, les variables définissant les états du système scolaire, le sexe, la langue, la localisation géographique, le régime d'étude, le statut de l'étudiant et l'âge.

Toutes ces variables sont importantes en ce sens que les sous-populations qu'elles définissent risquent de se comporter différemment, c'est-à-dire, d'avoir des coefficients de transition différents. L'exemple suivant montre cependant qu'il nous

faut garder un souci d'efficacité et de simplicité.

Si au 1er cycle universitaire on répartit la population selon le sexe, on obtient 96 X 2 = 192 états. La prise en compte de l'institution fréquentée en même temps que du sexe nous mène à 96 X 7 X 2 = 1344 états; et ainsi de suite.

On constate aisément qu'avec un minimum de variables on aboutit rapidement à un nombre important de possibilités. Accepter dès le départ un trop grand nombre de variables compromettrait, pensons-nous, le succès du travail. Le raffinement du modèle devra se faire par étape, et en étroite collaboration avec les informaticiens chargés d'écrire les programmes de traitement des données. Nous commencerons en ne retenant que les sous-populations les plus globales.

2) Définition de la variable temps

Il s'agit ici de définir la période de temps (ou l'étape) pour laquelle on calculera les coefficients de transition .

Pour les niveaux infra-collégiaux, la période de temps visée sera l'année scolaire et les flux seront calculés de 30 septembre en 30 septembre.

Aux niveaux collégial et universitaire, l'étape est la session. Il faudra donc calculer les flux entre chaque session, ce qui implique une cueillette des inscriptions pour chacune des sessions.

3) Données nécessaires au calcul des coefficients

Le calcul des coefficients de transition entre les étapes t et t+1 exige une classification détaillée des inscriptions à la date x du temps t et une compilation identique à la date x du temps t+1, ventilée selon la provenance, c'est-à-dire l'activité à la date x du temps t. La deuxième compilation fournit (en les distinguant) les immigrants dans le système (les nouveaux inscrits).

Le fichier individuel cumulatif avec code permanent permet la production d'une telle information(1).

L'absence de telles informations pour le passé nous conduira, comme on le verra maintenant, à estimer les flux passés.

Méthodologie des prévisions

Pour nos prévisions des populations scolaires, nous comptons utiliser des méthodes tout-à-fait traditionnelles en démo-

---

(1) Une question posée à l'individu à la cueillette x. t+1 sur son état au moment x.t permettrait également d'aboutir au même résultat.

graphie pour les perspectives de population totale (méthode des composantes). On multiplie un secteur de population au temps t par une matrice correspondant aux lois de mortalité et de fécondité de la population. On ajoute au résultat un vecteur donnant la population migrante nette qui est estimée empiriquement.

Dans le cas des perspectives de la population scolaire nous utilisons

1) un vecteur de la population scolaire répartie selon les états choisis
2) une matrice des coefficients de transition d'un état à un autre au cours d'une étape
3) un vecteur d'immigration correspondant aux entrées dans le système scolaire.

L'évolution des coefficients de transition sera prévue à partir de l'analyse des tendances passées[1], à l'aide de méthodes de régression, de moyennes mobiles, etc. En combinant diverses méthodes on pourra obtenir une fourchette des valeurs attendues.

---

(1) Même si nous n'avons pas des statistiques de flux dans le passé il est possible à l'aide d'autres indices, d'obtenir des estimations de la valeur des coefficients de transition. L'annexe III présente une tentative de découverte des liens entre ces indices.

Le problème des entrées est beaucoup plus complexe car on n'a pas d'indice permettant de les mesurer. Le B.S.Q. nous fournira la population âgée de 5 ans pour les années à venir, ce qui nous permettra d'estimer les entrées à la maternelle. Mais nous connaissons mal les entrées provenant de l'extérieur du système scolaire (immigrants provenant de l'extérieur du Québec, retours après abandon temporaire des études ("drop out", malades...). Il faudra faire ici des estimations empiriques.

Etapes à venir

Les étapes du travail à accomplir dans un avenir rapproché sont donc les suivantes:

1) Description détaillée des états du système scolaire.

2) Description du contenu des fichiers cumulatifs - Adéquation avec le système scolaire.

3) Analyse des tendances passées - prévisions préliminaires globales.

4) Expérimentation partielle de la mesure des flux à l'aide des fichiers cumulatifs.

Annexe I

S I G U

DESCRIPTION SOMMAIRE ET PRELIMINAIRE D'UN MODELE DE FLUX DES CLIENTELES ETUDIANTES

1.- Définition du champ des clientèles

Le modèle de flux analysera les mouvements des clientèles scolaires des niveaux

- secondaire,
- collégial, général et professionnel,
- universitaire du 1er cycle et du 2e cycle professionnel.

2.- Utilisation du modèle - Objectifs

Le modèle servira aux fins de prévisions de clientèles et de diplômés ainsi qu'aux fins d'études sociologiques sur les clientèles scolaires.

Remarque: La description qui suit ne s'applique qu'aux clientèles de niveau universitaire. Elle pourra être étendue aux autres niveaux d'enseignement après entente avec les responsables de ces niveaux au ministère de l'Education.

2.1 Prévisions

2.11 Prévisions de diplômés

Objectifs:
prévoir le nombre de diplômés, ventilé selon:

l'institution universitaire fréquentée,
le secteur linguistique,
le sexe,
le lieu d'origine (à préciser).

a) par programme selon la taxonomie du SIGU, pour une période de 3 ou 4 ans dépendant de la durée des études,

b) par secteur de programmes, pour une période de 2 années additionnelles,

c) pour l'ensemble des secteurs du 1er cycle (et du 2e cycle professionnel), pour une autre période de 5 années additionnelles.

2.12 Prévision des nouvelles clientèles

Objectifs:
prévoir le nombre d'entrants selon la ventilation donnée à l'article 2.11,

a) par programme ou secteur de programmes (à établir) pour une période de 2 ans,

b) pour l'ensemble des secteurs du 1er cycle (et du 2e cycle professionnel) pour une période additionnelle de 5 ans.

2.13 Prévision des clientèles en formation

Objectifs:
prévoir les clientèles universitaires selon la ventilation donnée à l'article 2.11 et selon l'année probable de graduation,

a) par programme ou secteur de programmes, (à établir) pour une période de 2 ans,

b) pour l'ensemble des secteurs du 1er cycle (et du 2e cycle professionnel), pour une période de 5 années additionnelles.

On tiendra compte dans ces prévisions non seulement des nouveaux entrants et des diplômés mais également des cas d'abandon et de changements de programmes et d'institutions.

Le modèle devra englober les clientèles dites à temps partiel et devra pouvoir s'accomoder de tous les régimes de promotion (par année, par semestre ou par matière).

2.2 Etudes sociologiques

L'exploitation du modèle devra rendre possibles des études sur:

- l'orientation et la motivation des étudiants,
- les échecs, les abandons et les changements de programmes et d'institutions,
- les pratiques d'admission et de promotion des institutions,
- la mobilité géographique des clientèles,

par l'étude des mouvements de clientèles analysés selon certaines caractéristiques individuelles comme:

- l'âge,
- le sexe,
- le groupe linguistique,
- le groupe socio-économique,
- l'origine géographique (à préciser),
- les succès scolaires antérieurs,
- l'état civil,
- les institutions de formation fréquentées.

3.- Elaboration du modèle de flux

Les responsables de l'élaboration du modèle de flux des clientèles devront:

- proposer
  - a) un modèle théorique de flux,
  - b) un système de prévision des clientèles,
  - c) un système d'analyse pour fins d'études sociologiques,
- s'assurer, de l'existence et de la validité des données qui alimenteront le modèle et, le cas échéant, spécifier la structure et le contenu des fichiers étudiants en fonction des exigences du modèle de flux.

- spécifier la précision des résultats obtenus par l'exploitation du modèle pour les fins indiquées et ce pour les valeurs totales et les variations annuelles de ces valeurs.

- faire une première exploitation du modèle avec des données réelles.

Annexe II

Coefficients de transition, matrices et chaînes de Markov. Application au système scolaire

1) Coefficients de transition

Dans un système où il y a plusieurs états[1] (a,b,...) et où on peut passer d'un état à un autre au cours d'une étape,[1] on appelle coefficient de transition la proportion de ceux qui étant en i au temps t, se retrouveront en un état quelconque j en t+1. On interprète parfois ces proportions (déterministes) comme des probabilités et on emprunte alors une approche stochastique.

Un coefficient de transition est nécessairement plus grand ou égal à 0, plus petit ou égal à 1, et la somme des coefficients de transition d'un état à tous les états (incluant l'état d'origine) est égale à 1. Remarquons que nous considérons une stagnation en a quand même comme une transition ou un passage de a à a.

On peut définir les coefficients de transition.

q_{ij} = fraction des éléments dans l'état i qui se retrouveront en j à l'étape suivante, i et j variant de 1 à n, n étant le nombre d'états.

---

(1) Nous suggérons l'utilisation des concepts état et étape pour décrire les niveaux et la durée.

on a

$$q_{ij} \geq 0, \leq 1 \quad (i,j = 1, n)$$

$$\sum_{j=1}^{n} q_{ij} = 1 \quad (i = 1,n)$$

2) Matrices des coefficients de transition.

On peut former des matrices avec l'ensemble des coefficients de transition.

On obtient la matrice A

$$A = \begin{bmatrix} q11 \ldots\ldots\ldots q1n \\ \cdot \quad\quad\quad\quad \cdot \\ \cdot \quad\quad\quad\quad \cdot \\ \cdot \quad\quad\quad\quad \cdot \\ \cdot \quad\quad\quad\quad \cdot \\ qn1 \ldots\ldots\ldots qnn \end{bmatrix}$$

Les lignes correspondent à des vecteurs de probabilité. Les coefficients sont tous plus grands ou égaux à 0 et la somme des coefficients d'une ligne est 1.

Comme la matrice est carrée (n colonnes et n lignes) et que toutes les lignes sont des vecteurs de probabilité, nous avons une matrice dite STOCHASTIQUE.

Avec un vecteur E de la "population" selon les n états à l'étape t, nous pouvons obtenir la répartition de la population en t+1 (s'il n'y a pas d'entrants provenant de l'extérieur) en le post multipliant (si E est un vecteur-ligne) par la matrice des coefficients de transition.

$$ {}_{t}E \times \Lambda = {}_{t+1}E $$

$$ ({}_{t}e_1, \ldots {}_{t}e_n) \begin{bmatrix} q11 \ldots\ldots\ldots q1n \\ \cdot \quad\quad\quad\quad \cdot \\ \cdot \quad\quad\quad\quad \cdot \\ \cdot \quad\quad\quad\quad \cdot \\ \cdot \quad\quad\quad\quad \cdot \\ qn1 \quad\quad\quad\quad qn,n \end{bmatrix} = ({}_{t+1}e \ , \ldots\ldots {}_{t+1}e_n) $$

Si E est un vecteur-colonne, nous devons le prémultiplier par la transposée de la matrice des coefficients de transition.

$$ \Lambda^{t} \times {}_{t}E = {}_{t+1}E \qquad (1) $$

Exemple: Un système scolaire comprend 3 états[1] (niveaux...nombre de crédits). On connaît les probabilités de passer d'un état à un autre au cours d'une étape (session ou année), la probabilité des finissants d'avoir un diplôme[2] et les probabilités d'abandonner après chaque état.[3] On suppose qu'il n'y a pas de retours et que les départs ont lieu à la fin de l'étape. Notre matrice est:

---

(1) 1, 2, 3

(2) d

(3) a, a_2, a_3

	1	2	3	d	a_1	a_2	a_3
1	.1	.8	0	0	.1	0	0
2	0	.1	.7	0	0	.2	0
3	0	0	.1	.8	0	0	.1
d	0	0	0	1	0	0	0
a_1	0	0	0	0	1	0	0
a_2	0	0	0	0	0	1	0
a_3	0	0	0	0	0	0	1

= A

Distribuons notre population au temps t

1 2 3 d $a_1 a_2 a_3$

t^E = (100,100,100,0,0,0,0)

A l'étape t+1 nous aurons

$t+1^E$ = (10,90,80,80,10,20,10) = t^E X A

Si les coefficients sont constants dans le temps, on peut connaître la répartition de la population en t+2 en postmultipliant $t+1^E$ par A ou encore en postmultipliant t^E par la matrice A au carré (A^2). $t+2^E = t+1^E \times A = t^E \times A^2$

et $t+n^E = t^E \times A^n$

on peut alors utiliser les chaînes de Markov.

Si les coefficients ne sont pas constants, il faut alors prévoir leur évolution future dans le temps. La matrice A doit alors être indicée t^A, $t+1^A$... On a dans ce cas;

$$t+1^E = t^E \times t^A$$
$$t+2^E = t+1^E \times t+1^A = t^E \times t^A \times t+1^A$$
$$t+n^E = t^E \times t^A \times t+1^A \times \ldots \times t+n-1^A$$

Ce sera probablement la méthode utilisée pour les projections des populations scolaire. Il faudra évaluer les matrices de transition du futur, y intégrer les coefficients de sortie. Mais il faut aussi estimer les vecteurs des entrants ou des immigrants dans le système. Il faudra faire un compromis entre l'analyse et l'efficacité. On devra calculer des coefficients de transition pour les groupes les plus efficients et significatifs. Le flair et l'expérience jouent alors un rôle important. Pour décrire l'histoire d'une cohorte quelconque, de façon longitudinale, il faut utiliser une nouvelle matrice à chaque étape. Il est aussi possible de décrire à chaque étape l'histoire d'une cohorte fictive (analyse transversale). On rejoint alors la situation décrite au paragraphe suivant.

3) Matrice des coefficients de transition constante.

Si les coefficients de transition sont constants on se trouve en situation stable et l'étude longitudinale donne les mêmes résultats que l'étude transversale.

Nous pouvons utiliser certaines propriétés des matrices stochastiques et obtenir facilement une description très détaillée de la vie de cohortes fictives.

Il nous faut considérer la matrice stochastique comme la matrice de transition d'une chaîne de Markov finie. Il nous faut aussi considérer les coefficients comme des probabilités et supposer que le processus est markovien ou à mémoire courte (la probabilité de passer de i à j ne dépend pas des états antérieurs à i)

La matrice A présente des états absorbants (d'où on ne sort pas; une fois diplômé, par exemple, on ne revient pas aux états antérieurs). Il est donc intéressant de diviser notre matrice en quatre sous-matrices.

$$Q = \begin{bmatrix} .1 & .8 & 0 \\ 0 & .1 & .7 \\ 0 & 0 & .1 \end{bmatrix} \qquad R = \begin{bmatrix} 0 & .1 & 0 & 0 \\ 0 & 0 & .2 & 0 \\ .8 & 0 & 0 & .1 \end{bmatrix}$$

$$0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

La matrice Q nous donne les coefficients de passage entre états non-absorbants, alors que la matrice R nous donne les coefficients de passage d'un état non-absorbant à un état absorbant. La matrice O est une matrice nulle et la matrice I une matrice identité.

Exemple:

Il est intéressant de présenter un exemple concret d'utilisation des matrices de coefficients. La matrice W (définie ci-dessous) donne le temps moyen (le nombre d'étapes) passé dans l'état j par les élèves entrant dans l'état i.

$$W = (I-Q)^{-1}$$

W est l'inverse de la différence entre la matrice identité et la matrice des coefficients Q.

Avec notre exemple (page 19) nous avons

$$(I-Q) = \begin{bmatrix} .9 & -.8 & 0 \\ 0 & .9 & -.7 \\ 0 & 0 & .9 \end{bmatrix}$$

Le déterminant de $(I-Q) = .729$

La matrice des cofacteurs de $(I-Q) = \begin{bmatrix} .81 & .72 & .56 \\ 0 & .81 & .63 \\ 0 & 0 & .81 \end{bmatrix} = MC$

On multiplie MC par le scalaire 1/déterminant

$$W = (I-Q)^{-1} = \begin{bmatrix} 1.11 & .987 & .768 \\ 0 & 1.11 & .864 \\ 0 & 0 & 1.11 \end{bmatrix}$$

Les nouveaux entrants à l'état 1 passeront donc 1.11 étapes en 1, .99 étapes en 2 et .77 étapes en 3 soit en tout 2.76 étapes dans le système scolaire. On obtient ainsi une espérance de vie scolaire, et ce, pour chaque état à l'entrée, qu'on peut repartir entre les différents états du système. (Rappelons que nous faisons l'hypothèse que les départs ont lieu à la fin d'une étape).

Cet exemple illustre bien l'intérêt de l'utilisation de telles matrices.

Nous pouvons étudier le dynamisme interne des mouvements et comprendre les implications des coefficients de transition. Il s'agit de simuler l'expérience scolaire grâce à une cohorte fictive.

Il est aussi possible de décrire la vie d'une cohorte réelle quand les matrices varient mais les calculs sont alors beaucoup plus compliqués et il faut les recommencer pour chaque

cohorte. Par exemple si l'on veut connaître combien d'étapes les entrants en j passeront en i la formule est:

$$W_{ij} = q_{ij}^{(o)} + q_{ij}^{(1)} + q_{ij}^{2} + \ldots \quad (i,j=1,n)$$

$q_{ij}^{(o)} = 0$ (si $i \neq j$), $= 1$ (si $i=j$)

$q_{ij}^{(1)}$ = élément i,j de la matrice t^Q

$q_{ij}^{(2)}$ = élément i,j de la matrice $t^Q \times t+1^Q$

Après quelques étapes qij a une valeur négligeable.

Remarquons que nous pouvons utiliser cette formule avec une matrice constante et obtenir les mêmes résultats qu'avec la matrice W mais le travail est plus onéreux.

Annexe III

Relations entre quelques indices utilisés en démographie scolaire: les taux de scolarisation, les taux de passage et les coefficients de transition

1. Définitions usuelles

1.1 Taux de scolarisation (T.S.)

On l'utilise parfois comme un indice très brut, soit par exemple le nombre d'étudiants d'un pays sur la population âgée de 5 à 29 ans. C'est évidemment alors un indice très peu raffiné, fortement influencé par des phénomènes de structure. Le T.S. par âge est beaucoup plus satisfaisant. La population scolaire d'âge x est rapportée à la population totale d'âge x à une date donnée. On trouve aussi des T.S. par âge/niveau et par âge/classe; les étudiants sont alors répartis selon les classes et/ou les niveaux. La somme de ces taux pour un âge donne évidemment le taux de scolarisation par âge. Les T.S. sont donc essentiellement des indices de génération.

1.2 Coefficient de transition (C.T.)

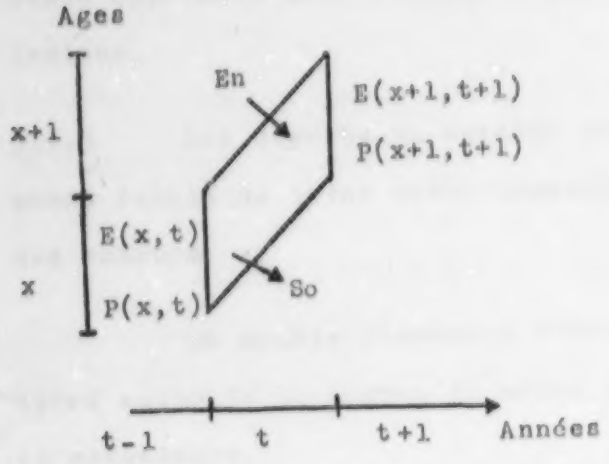
L'ensemble des individus vivant un événement a constitue une cohorte. Au bout d'un certain temps ces individus vivent d'autres événements (b, c, d...) ou restent en a. Les coefficients de transition sont les probabilités de passage de a à a, b, c, d... La somme des coefficients est égale à 1.

En démographie scolaire les C.T. les plus utilisés sont les coefficients de promotion (proportion des étudiants de la classe x en l'année scolaire t qui seront en classe x+1 en l'année scolaire suivante), les coefficients de répétition et les coefficients de départs.

1.3 Taux de passage (T.P.)

Ce taux rapporte les étudiants de la classe x+1 de l'année t+1 aux étudiants de la classe x de l'année t. Il s'agit donc ici de cohortes "apparentes". (Les individus au numérateur ne se retrouvent pas nécessairement au dénominateur).

2. Symboles utilisés



x = âge

t = année

P = population totale

E = population scolarisée

En= entrées

So= sorties

3. Modèle

3.1 Afin d'étudier les relations entre les indices il faut se donner un modèle du système scolaire, faire certaines hypothèses simplificatrices qui permettent de faire varier un paramètre à la fois.

3.1.1 Il n'y a pas de redoublement et tous les élèves entrent dans le système à un âge donné. (Système de progrès continu). L'étude par classe sera donc égale à l'étude par génération.

3.1.2 Il n'y a pas de décès. Dans un premier temps nous supposerons qu'il n'y a pas de migration puis nous lèverons cette hypothèse pour étudier l'effet des migrations sur nos indices.

3.1.3 Les départs du système sont définitifs. Il sera aussi facile de lever cette hypothèse pour voir l'influence des retours.

Le modèle ressemble d'assez près à la structure scolaire actuelle du Québec du moins pour les niveaux élémentaire et secondaire.

3.2 Modèle sans migration

Les indices seront donc définis:

$$T.S.\ (x+1,\ t+1) = \frac{E\ (x+1,\ t+1)}{P\ (x+1,\ t+1)}$$

$$T.S.\ (x,\ t) = \frac{E\ (x,\ t)}{P\ (x,\ t)}$$

Nous n'utiliserons qu'un coefficient de transition que nous appellerons le coefficient de passage (C.P.). Etant donné le système à progrès continu nous définirons le C.P. la proportion des individus scolarisés au temps t qui seront toujours scolarisés au temps t+1. Comme il n'y a pas de migration, $C.P. = \frac{E\ (x+1,\ t+1)}{E\ (x,\ t)}$

Avec ce modèle sans migration, et par génération, le T.P. est égal au C.P. $T.P. = \frac{E\ (x+1,\ t+1)}{E\ (x,\ t)}$

Il est intéressant de mesurer le rapport entre les T.S. Comme par définition, P (x+1,t+1) = P (x,t), ce rapport (R.T.S.) est égal à C.P. et à T.P.

Utilisons un exemple empirique pour illustrer les relations entre les indices.

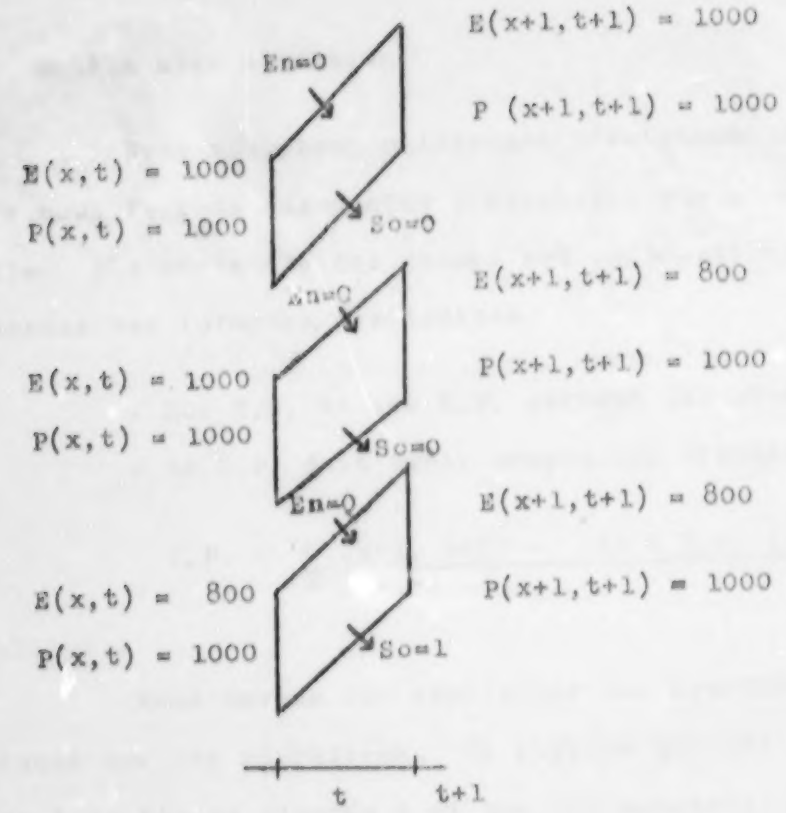


Tableau 1

T.S. (x,t)	T.S.(x+1, t+1)	R.T.S.	C.P.	T.P.
1.00	1.00	1.00	1.00	1.00
1.00	.80	.80	.80	.80
.80	.80	1.00	1.00	1.00

Les indices R.T.S., C.P. et T.P., ne nous indiquent rien sur le niveau de la scolarisation à l'année t ou t+1 mais indiquent seulement le succès du système à relever les étudiants.

3.3 Modèle avec migration

Nous supposons maintenant l'existence de migrations mais nous faisons quand même l'hypothèse que la mortalité est nulle. (La mortalité des jeunes est négligeable) Il nous faut réécrire les formules des indices.

- Les T.S. et les T.P. gardent les mêmes définitions
- Le C.P. doit tenir compte des migrations

$$C.P. = \frac{E\ (x+1,\ t+1) - \text{En X T.S. } (x+1,\ t+1)}{E\ (x,\ t)}$$

Nous devons ici expliciter les hypothèses que nous faisons sur les migrations. On suppose que les entrées ont lieu à la fin de l'année t et que les entrants auront un T.S. égal à celui de l'ensemble de la population au temps t+1. D'autre part il y a également 'ndépendance des probabilités de sortie et de promotion. Les émigrants ont le même taux de scolarisation que la population au temps t.

Nous supposons toujours que les abandons sont définitifs.

Reprenons le modèle empirique pour expérimenter les effets des migrations.

Nous utiliserons trois niveaux d'immigration et d'émigration.

En = 0, 100, 200; So = 0, 100, 200

Pour calculer les T.P. et les C.P. il nous faut connaître: P(x,t) (ici égal à 1000)

T.S.(x,t), T.S.(x+1,t+1), En et So

$$T.P. = \frac{(P(x,t) - So+En) \times T.S.(x+1,t+1)}{P(x,t) \times T.S.(x,t)}$$

$$C.P. = \frac{(P(x,t) - So+En) \times T.S.(x+1,t+1) - En \times T.S.(x+1,t+1)}{P(x,t) \times T.S.(x,t)}$$

Tableau 2

T.S. (x,t)	T.S. (x+1,t+1)	R.T.S.	En / So	200 / 200	100 / 200	0 / 200	200 / 100	100 / 100	0 / 100	200 / 0	100 / 0	0 / 0
1.0	1.0	1.0	T.P.	1.0	.9	.8	1.1	1.0	.9	1.2	1.1	1.0
			C.P.	.8	.8	.8	.9	.9	.9	1.0	1.0	1.0
1.0	.8	.8	T.P.	.80	.72	.64	.88	.80	.72	.96	.88	.80
			C.P.	.64	.64	.64	.72	.72	.72	.80	.80	.80
.8	.8	1.0	T.P.	1.0	.9	.8	1.1	1.0	.9	1.2	1.1	1.0
			C.P.	.8	.8	.8	.9	.9	.9	1.0	1.0	1.0

Quand la migration est nulle, on retrouve évidemment les résultats du tableau 1.

Les C.P. varient en raison inverse avec l'imigration (les étudiants immigrants ne se trouvent pas au numérateur) et n'indiquent pas le niveau de la scolarisation mais seulement le

succès des écoles à retenir leur clientèle (les C.P. varient non pas avec les T.S. mais avec le rapport entre deux T.S. successifs (l'émigration étant tenue constante)).

Alors que le C.P. nous renseigne sur une cohorte réelle, le T.P. décrit une cohorte apparente.

Le T.P. ne nous renseigne pas sur le niveau de la scolarisation ni sur le succès du système scolaire à retenir ses étudiants, il varie en relation directe avec la migration nette et en relation directe avec le rapport entre les taux de scolarisation. C'est donc l'indice le plus grossier et le plus difficile à interpréter.

Pour décrire la scolarisation en analyse longitudinale pour une génération donnée, il est souhaitable de connaître à la fois les T.S. et les C.P. surtout si l'émigration est importante. Dans un système à progrès continu, où par surcroît la scolarité est obligatoire, comme au Québec à l'élémentaire et au secondaire (grosso modo), les R.T.S. seront égaux à ou très près de l'unité mais les C.P. nous renseigneront sur l'émigration ou encore si on connaît la probabilité d'émigrer, on pourra évaluer le C.P.

Il n'est donc pas très utile d'étudier les flux étudiants pour les niveaux élémentaire et secondaire quand la scolarité est obligatoire et le progrès continu. Mais il y a

quand même, en pratique, des retards scolaires et des migrations. L'étude des flux à l'élémentaire et au secondaire sera intéressante pour la description des migrations, des orientations générale et professionnelle et des retards au secondaire, et aidera ainsi à prévoir plus exactement l'évolution des populations scolaires au niveau régional. Ce serait aussi intéressant de connaître certains transferts d'étudiants (par exemple: changement de la langue d'enseignement, changement de programme au secondaire...).

Si la scolarisation n'est pas obligatoire, il faudra connaître à la fois le R.T.S. et la probabilité d'émigrer (P.S.) pour évaluer le C.P.

P.S. = So/P(x,t)

C.P. = R.T.S. X (1-P.S.)

La connaissance des C.P. est ici beaucoup plus pertinente que lorsque la scolarité est obligatoire. Et comme la mesure des émigrants est toujours assez précaire, il est plus prudent de mesurer directement les C.P. que d'utiliser le recours précaire à l'estimation des émigrants. Comme, lorsque la scolarité n'est pas obligatoire, le progrès n'est pas continu, nous définissons le C.P. comme la probabilité pour les étudiants en t d'être encore scolarisés en t+1.

3.4 Modèle avec retours

Les retours ne sont possibles que lorsque la scolarisation n'est pas obligatoire et on ne peut plus alors parler de progrès continu.

L'étude par génération ou par âge perd alors un peu de sa valeur exclusive mais reste très intéressante. Si on ne tient pas compte des classes ou des étapes, on peut quand même décrire l'expérience scolaire d'une génération. Pour reprendre la terminologie de J. Légaré on s'intéresse à l'efficience et non pas à l'efficacité de la scolarisation pour une génération (on saura combien d'années elle passera sur les bancs d'école sans se préoccuper des diplômes reçus...)

Comme les retours sont possibles, il nous faut donc refaire les hypothèses conduisant à la définition du C.P.

On peut supposer (hypothèse 1) que tous les étudiants de l'année t qui seront scolarisés en t+1 le seront avant qu'un individu ne fréquentant pas l'école en t ne devienne scolarisé en t+1 (retour).

Le moyen de calculer le C.P. (1) est assez complexe.

$A = (P(x,t) - So) \times T.S.(x+1,t+1)$

$B = (P(x,t) - So) \times T.S.(x,t)$

si $A \geqslant B$, $C.P._{(1)} = B / P(x,t) \times T.S.(x,t)$

si $A \leqslant B$, $C.P._{(1)} = A / P(x,t) \times T.S.(x,t)$

On peut encore supposer (hypothèse 2) qu'il y a indépendance entre la scolarisation en t et en t+1 (ceux qui n'étudient pas en t ont autant de chances d'être scolarisés en t+1 que les autres).

La formule du C.P. (2) est:

(1)

$$C.P. = \frac{(P(x,t) \times T.S.(x,t) - So \times T.S.(x,t)) \times T.S.(x+1,t+1)}{P(x,t) \times T.S.(x,t)}$$

Tableau 3

T.S. (x,t)	T.S. (x+1,t+1)	R.T.S.	En / So	200 / 200	100 / 200	0 / 200	200 / 100	100 / 100	0 / 100	200 / 0	100 / 0	0 / 0
.8	.8	1.0	T.P	1.0	.9	.8	1.1	1.0	.9	1.2	1.1	1.0
			C.P(1)	.8	.8	.8	.9	.9	.9	1.0	1.0	1.0
			C.P(2)	.64	.64	.64	.72	.72	.72	.8	.8	.8
.8	.9	1.125	T.P	1.125	1.01	.90	1.24	1.125	1.01	1.36	1.24	1.125
			C.P(1)	.8	.8	.8	.9	.9	.9	1.0	1.0	1.0
			C.P(2)	.72	.72	.72	.81	.81	.81	.9	.9	.9

Si les retours sont possibles, il ne faudra donc pas s'étonner de trouver parfois des T.S. qui augmentent dans le temps pour une même génération.

Même si le T.S. ne varie pas entre t et t+1, la présence d'étudiants en t+1 qui n'étaient pas scolarisés en t implique que le C.P. des étudiants de t est plus faible. Il est probable que l'hypothèse conduisant au calcul du C.P.(2) est trop généreuse envers les chances de retour à l'école des jeunes qui ont déjà abandonné leurs études. La "réalité" doit se situer quelque part

1) on suppose toujours P(x,t) = 1000

entre les 2 hypothèses. Il est en effet très peu probable (mais non impossible) que les jeunes qui ont quitté la vie d'étudiant aient plus de chances d'y retourner que les étudiants d'y rester.

On voit donc que l'existence de retours complique et même détruit la relation univoque entre les T.S. et les C.P.

Avec C.P.(1) nous avions (page 29):

$C.P._{(1)} = R.T.S. \times (1 - (So/P(x,t))$

Cette relation n'est plus vraie quand $T.S.(x+1,t+1) > T.S.(x,t)$

Le R.T.S. et le T.P. peuvent être >1 mais pas le C.P.

Avec C.P.(2) nous avons:

$C.P._{(2)} = T.S.(x+1,t+1) \times (1 - (So/P(x,t))$

donc quand $R.T.S. \leq 1$

$$\frac{C.P.(1)}{C.P.(2)} = \frac{T.S.(x+1,t+1)}{T.S.(x,t)} \times \frac{1}{T.S.(x+1,t+1)}$$

$$C.P.(2) = C.P.(1) \times T.S.(x,t)$$

Quand $R.T.S. \geq 1$ nous avons toujours

$$C.P.(2) = T.S.(x+1,t+1) \times (1 - (So/P(x,t))$$

Mais $$C.P.(1) = \frac{C.P.(2)}{T.S.(x+1,t+1)}$$

$$= 1 - (So/P(x,t))$$

Le C.P.(1) quand R.T.S. ≥ 1 ne varie qu'avec l'émigration alors que lorsque R.T.S. < 1 il varie également avec le R.T.S. (Quand R.T.S. = 1, T.S.(x,t) = T.S.(x+1,t+1) et les deux relations sont vérifiées).

L'écart entre C.P.(1) et C.P.(2) forme la fourchette des valeurs possibles du C.P. étant donné T.S.(x,t), T.S.(x+1,t+1) et les émigrations. Dans un modèle sans retours, la connaissance de ces variables nous donnait la valeur du C.P. D'autre part, dans l'un comme dans l'autre modèle, la connaissance du C.P. ne nous renseigne en rien sur les T.S. Si on connaît l'émigration, dans le modèle sans retour, on peut déduire le R.T.S. mais non la valeur de T.S.(x,t) ou de T.S.(x+1,t+1) alors que dans le modèle avec retours, on ne peut même obtenir le R.T.S. de façon exacte.

L'étude de la scolarisation par génération seulement devient ainsi moins pertinente et on sent le besoin de combiner les variables âge et classe (ou étape), pour obtenir une meilleure description de la vie scolaire. Avec des C.P. par classe et âge on peut passer aux C.P. par âge et aux C.P. par classe.

4) Conclusion

4.1 Modèles des systèmes scolaires possibles

				Modèle
Scolarité obligatoire	progrès continu	Sans retours	Sans migration	1
			Avec migration	2
		X X		
	classes étapes	Sans retours	Sans migration	3
			Avec migration	4
		X X		
Scolarité non-obligatoire	X X			
	classes étapes	Sans retours	Sans migration	5
			Avec migration	6
		Avec retours	Sans migration	7
			Avec migration	8

4.2 Quand la scolarité est obligatoire, il ne peut y avoir de retours (puisqu'il n'y a pas en théorie de départs) et le système de progrès continu (la classe est déterminée par l'âge) est incompatible en pratique avec une scolarisation non-obligatoire.

Le système scolaire du Québec se rapproche du modèle 2 pour l'élémentaire et une bonne partie du secondaire. On passe ensuite au modèle 8: après 15 ans, la scolarisation n'est plus

obligatoire, il y a des étapes à franchir, des retours possibles et il faut tenir compte des migrations.

Un système à progrès continu (qui implique la scolarisation obligatoire et l'absence des retours) facilite la tâche du démographe puisque l'étude par génération correspond étroitement à l'étude par cohortes.

Par contre, quand il n'y a pas de progrès continu, l'utilisation des cohortes (population d'une classe par exemple) devient conseillée car il nous faut maintenant distinguer l'efficacité et l'efficience du système. Comme les liens entre les C.P. et les T.S. ne sont pas univoques, il nous faudra, pour l'étude de données observées, utiliser les deux indices. On pourrait à la rigueur, en faisant un peu de cuisine démographique (utiliser la méthode des moyennes pondérées, faire des hypothèses sur les retours, etc.) obtenir une certaine estimation des T.S. à partir des C.P.

Nous avons tenté d'éclairer les liens entre divers indices utilisés en démographie scolaire, en tenant compte surtout de l'analyse par génération. Quand il n'y a pas de progrès continu et qu'on admet les retours, les relations entre les indices sont loin d'être univoques.

Les modèles que nous avons utilisés sont beaucoup plus complexes que les modèles employés généralement car nous avons tenu compte des migrations et des retours. Par contre, nous avons surtout concentré nos efforts à l'étude par génération.